DOCUMENT RESUME

ED 242 774 TM 840 191

AUTHOR Meyers, H. W.
TITLE Disaggregation of Student Outcome Data in Pre-Assessment: More than a Snapshot.
PUB DATE 26 Apr 84
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Ability; *Academic Education; Data Analysis; Educational Environment; *Educational Trends; Instructional Improvement; Instructional Program Divisions; Secondary Education; *Social Differences; *Standardized Tests; *Test Results
IDENTIFIERS Aggregation (Data); Comprehensive Tests of Basic Skills; *Difference Scores; *Disaggregation (Data)

ABSTRACT

Disaggregation of standardized test data is essential in determining the differential effects of the school's climate and program upon groups of students formed by social class. In a recent secondary school study comparing the outcomes of two approaches to school assessment (the Connecticut School Effectiveness Survey vs. The National Study of School Evaluation), standardized Comprehensive Test of Basic Skills test data were disaggregated not only by social class, but also by class year, sex, program of study, and ability levels. In addition, data were analyzed for trends across three years of testing for each student. Findings included some surprising trends within the whole of the student population when disaggregated by the above variables, while holding the effects of curriculum (class year) constant. Results suggest that the "snapshot" approach in disaggregating test data by social class may mask the condition that the school may be improving or deterioriating on one or more of the standard indicators of pupil performance. Since patterns of pupil performance are often correlated with teacher attitudes about school climate, program and students, the more detailed the analysis of test data the more useful it may become in driving school improvement. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED242774

Disaggregation of Student Outcome Data in Pre-Assessment: More Than A Snapshot

H.W. Meyers, Ph.D.
Center for Evaluation & Policy Research
The University of Vermont

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. W. Meyers

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 840 181

Paper presented at the American Educational Research Association, April 26, 1984, New Orleans, Louisiana

ABSTRACT

Disaggregation of Student Outcome Data in Pre-Assessment:

More Than A Snapshot

Disaggregation of standardized test data is essential in determining the differential effects of the school's climate and program upon groups of students formed by social class.[1] In a recent secondary school study comparing the outcomes of two approaches to school assessment (the Connecticut School Effectiveness Survey vs. The National Study of School Evaluation)[2] standardized CTBS test data were disaggregated not only by social class, but also by class year, sex, program of study and ability levels. In addition, data were analyzed for trends across three years of testing for each student. Findings included some surprising trends within the whole of the student population when disaggregated by the above variables, while holding the effects of curriculum (class year) constant. Results suggest that the "snapshot" approach in disaggregating test data by social class may mask the condition that the school may be improving or deterioriating on one or more of the standard indicators of pupil performance. Since patterns of pupil performance are often correlated with teacher attitudes about school climate, program and students, the more detailed the analysis of test data the more useful it may become in driving school improvement.

H.W. Meyers, Ph.D.
Center for Evaluation & Policy Research
The University of Vermont

---

[1]Pecheone, R. & Shoemaker, J. "The Use of Multiple Data Sources to Assess School Effectiveness." Paper presented at the Annual Conference of the American Educational Research Association, New York, 1982. p.6.

[2]Buttram, J. & Carlson, R. The Implications of Effective Schools Research to the National Study of School Evaluation (NSSE). In process, 1984.

- INTRODUCTION -

The application of school effectiveness evaluation strategies and findings to local secondary school improvement efforts had experienced some growth during the years 1980 through 1982. (Miles, et.al. 1983). Miles and his colleagues identified some 39 effective schools programs in 23 states, and these programs, by conservative estimate, represented implementation in about 1768 school buildings. Further, about 41% of the schools involved were secondary schools (p. 9).

The use of achievement data as the outcome measure for defining relative school effectiveness is generally accepted (Neufield, et.al. 1983, p. 16, 19) but not without some discussion and discontent among secondary administrators and teachers. The issues presented in the literature range from the inherent difficulties with using standardized tests as outcome measures, (Madaus, et.al. 1980) to the narrowness which the achievement outcome implies, to the effects of tracking on standardized test score distributions. (Neufield, p. 19). Added to the foregoing set of issues is the question of educational equity, defined by Edmonds (1982, p. 6) as: "... the opportunity for all students to advance to the next level of schooling."

This definition, in turn has led to the formulation of criteria for judging an effective school and to the identification of grouping variables by which analysis of achievement data are conducted. (Edmonds, 1982, p.6).

That the goal of using achievement scores to measure the relative effectiveness of school has not won universal acceptance is documented by Miles (1983, p.35) wherein his assessment of 35 school improvement programs indicated that 49 percent of the programs were using student achievement as an outcome measure. Limited acceptance is rationalized by Neufield and her colleagues for the reasons cited above (Farrar, 1983, p.27); and by Edmonds (1982, p.11) as indifference on the part of many schools to the questions of equity.

To school improvement programs designed around the five characteristics identified by Edmonds (1982, Airlie House) however, "the use of measures of pupil achievement as the basis for program evaluation" is a common feature. (Edmonds, 1982). Having accepted the notion that evaluation should at least include test scores as outcome measures, the next question addressed is ordinarily one of design, or how to array the scores and what school or background variables ought to be controlled. Here, a variety of statistical procedures are found. Early strategies, included regression analysis which focused on the "properties of the regression line, the percentage of variation explained (R^2), and the coefficients of the regressor variables." (Klitgaard & Hall, 1973, p. 16). While few effective schools were found by this method, attention to the residuals of the regression line produced the identification of "outlier" schools (p. 69-72). When schools of this type were studied in more detail (Edmonds, 1978) they helped lay the foundation for correlational research on factors associated with school effectiveness.

Correlational strategies employed in the search for instructionally effective schools focused on relationships of variables between and among schools. School improvement assessments, on the other hand, focus on pupil performance within schools. Each assessment, therefore, is an "N of 1" study and hence, descriptive statistics are both appropriate and probably better understood than more elegant inferential techniques by teachers, administrators and parents. For example, the New York "SIP" project typically reports its findings in terms of percentile ranks and percentages performing at grade level (Oratz, et. al., 1981, p. 11-18). RISE reports its results (or presumed effects) in terms of percentage of those students performing in the high and average categories on the Metropolitan Achievement Tests over a five year period. (McCormack et. al., Educational Leadership, Dec., 1982).

Pre-Assessment

It is important, at this juncture, to distinguish between the reporting of results of the effects of school improvement programs and the pre-assessment process

which schools engage in prior to planning for school improvement. The former process encourages the tracking and sorting of achievement data from inception to evaluation. The latter process, however, is often done as a "snapshot" of school achievement taken in order to provide a baseline of the achievement patterns among groups of students formed by social class. The Connecticut school improvement project recommends the disaggregation of achievement data by social class on the basis of percentile ranks and identifies the 30th percentile as the criterion for minimum mastery on a norm referenced standardized test of achievement. (Connecticut State Department of Education, Draft, "The Connecticut School Effectiveness Project: Development and Assessment", 1981, p. 19-20). The Connecticut guidelines, however, (Pecheone and Shoemaker, 1982) are silent on the use of longitudinal data. This lack of attention to longer term patterns of achievement within schools may seem strange in light of Shaycoft's (1967) analysis of the Project Talent data, wherein she found differences in growth rates of achievement patterns among schools.* But it is understandable given the finite resources available to schools to do the pre-assessment and the lack of sufficient test data over a period of years in many schools.

Neufield (1983, p.19) cites the Rutter et.al. (1979) study as an example of one longitudinal study: "... they used a longitudinal approach, studying their schools over a period of five years rather than taking a snapshot of effectiveness at a single point in time. If variables remain important and associated with outcomes over time, they will provide more credence to those thinking about creating them in their own schools."

If the purpose of achievement-outcome preassessment in the context of school improvement programs is to provide a school with reliable information concerning the patterns of achievement of its students disaggregated by social class (and any number of other independent variables) it follows that the pre-analysis is best done over time.

---

*See also Acland, Henry D. "School Effects: An Analysis of New York Elementary Schools," unpublished, and Klitgaard and Hall, p. 62-65.

Setting and Purpose

In 1983, NSSE (the National Study for School Evaluation) awarded a grant to Drs. Joan Buttram and Robert Carlson for a comparative study of the processes and data produced by a regional accreditation process (the New England region) and the Connecticut School Effectiveness Assessment process.** A rural high school of approximately 500 students, having recently completed regional accreditation was chosen as the site for the study. (Buttram and Carlson, 1984).

The major preassessment question concerning student achievement to be answered was the following: How do groups of students, who are not eligible for special education services, formed by the various measures of social class, grade, sex, and academic track perform on standardized measures of achievement, over time?

Method

Reading and Mathematics scores on the Comprehensive Test of Basic Skills (CTBS) were obtained from school records for 455 students during the years 1981, 1982 and 1983, for all four grades (9-12) at the high school. These scores were then transformed to normal curve equivalents (NCE) and along with other variables identified above, entered as data for processing with the Statistical Package for the Social Sciences (SPSS). Statistics used in the presentation of findings to school staff were mainly descriptive except in cases where small cell sizes indicated the need for evaluating chance variation between groups. In those instances, the chi-square was employed.

Results

The distribution of high school students on each of the grouping variables in Figure 1 offers few surprises in the context of other Vermont schools studied to date.

**As modified by the Center for Evaluation and Policy Research, University of Vermont.

Figure 1

Variable	N
Sex:	M = 220
	F = 235
Grade:	9 = 112
in 1983	10 = 126
	11 = 96
	12 = 121
Spec. Educ.:	Eligible = 29
	N/Eligible = 426

Variable	N
SES (counselor):	Low = 84
	Other = 371
Lunch:	Yes = 104
	No = 351
Curriculum:	College = 179
	Business = 51
	General/ Vocational = 225

One minor surprise was the fact that only about 7 percent of those tested were identified as eligible for special education and hence removed from further analyses. When these students are added to other more severely handicapped students who were not tested with CTBS however, the percentage is an expected 13 percent, and the remaining 5 percent of missing data were distributed across grades.

Following a typical snapshot approach, three years of test data for reading and mathematics scores shows little variation across three years of data for the total school population. (see Figure 2)

Figure 2

1981		1982		1983	
Reading	NCE = 46		NCE = 50.9		NCE = 51
	%ile = 42		%ile = 52		%ile = 52
Mathematics	NCE = 48	Mathematics	NCE = 51.8	Mathematics	NCE = 54.7
	%ile = 48		%ile = 54		%ile = 60
Total Battery	NCE = 43	Total Battery	NCE = 49.7	Total Battery	NCE = 50.7
	%ile = 36		%ile = 48		%ile = 52

When data were disaggregated by social class** (but across grades) in 1983 the following pattern emerged:

Figure 3

Comparison of all Students Who Fall Above and Below the Thirtieth Percentile on CTBS Total Battery, 1983

N = 377*

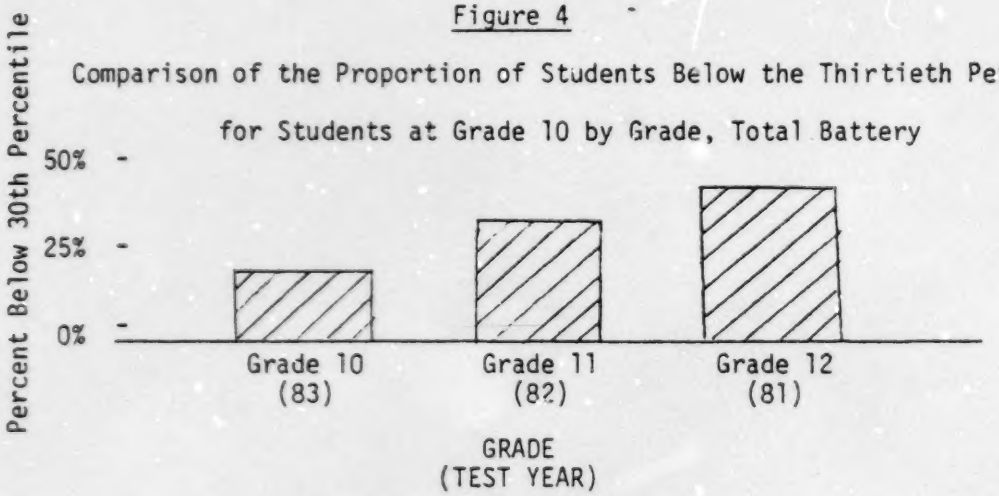
	Free or Reduced Lunch	Other
Above 30th Percentile	47 (47%)	213 (74%)
Below 30th Percentile	52 (53%)	75 (26%)

* Does not include special education or random missing scores.

Here again, given the patterns of achievement in other rural schools studied in Vermont, the distributions are not surprising. At this point in the analysis however, the opportunity was present to disaggregate the data by grade while holding test year constant, thus comparing the achievement of the current senior class when members of that class were sophomores with current sophmores, etc., the pattern of proportions which emerged across the classes is indicated in Figure 4.

Figure 4

Comparison of the Proportion of Students Below the Thirtieth Percentile for Students at Grade 10 by Grade, Total Battery



** Free and reduced lunch vs other (see later discussion on the correlation between lunch and staff judgments).

In holding the effects of grade constant it appeared that overall student performance improved with each successive class.

The observation of these differences in proportions below the 30th percentile by grade suggested that other differences in proportions above and below mastery in groups formed by SES, Sex and tracking might likewise be affected by grade differences. It should be noted here that disaggregation on the basis of SES and grade for the same calendar year may present descriptive statistics which may give the false impression of an increasingly effective school, i.e. a school where in the effect of curriculum acts in such a way as to reduce the proportion of low income children below minimum mastery over time. Many students (nearly 15 percent of the 1983 senior class) who were below the 30th percentile in grade 10 in 1981 however were no longer in school by grade 12.

Disaggregating the data by SEX and Grade produced interesting results. In reading and total battery scores females tended to out perform their male counterparts at each grade level, though the strength of this performance was greatest in grades 9 and 10. In mathematics, however, the picture was somewhat different. By following cohorts of students over a three year period it was observed that females lost nearly a standard deviation advantage between grades 10 and 12.

Tracking students placed in college bound, business and vocational tracks over a three year period showed negligible movement of students between and among curricular tracks. In addition, the proportion of students placed in each track by grade were similar across grades. Finally, the general/vocational track was a good predictor of low achievement.

Discussion

Few conclusions concerning the causes of variation in student performance may be drawn from the analysis of test data in this example of a pre-assessment conducted in one rural high school. Shoemaker and Pecheone's advice that presentation of school data should be descriptive rather than prescriptive, 1982 p. 14, and Klitgaard and Hall's, p. 65, notion that student outcome data be (dis)aggregated in a variety of

ways seem particularly relevant here. From these data, however, it can be said that the way in which data are disaggregated and portrayed may lead school planners to different conclusions about the nature of the patterns of achievement in their schools and hence to different strategies for school improvement.

If, for example, in this study only 1983 test data had been disaggregated by social class measures the picture presented would have been one of an ineffective school, one in which the proportion of low income children who achieve minimum mastery is considerably less than the proportion of other children who achieve mastery at or above the 30th percentile. In addition, teachers would have been presented with findings that suggested that the proportion of low income students who fell below the 30th percentile was declining in each grade. In fact however, when dropouts were accounted for the proportions of low income to other students below the 30th percentile remained relatively constant while the proportion of all students below the 30th percentile declined. In sum, the gap between low income and other students was actually widening rather than narrowing. (see Figure 5).

Figure 5

Proportion of Students Gaining Above the Thirtieth Percentile

Between 1981 and 1983

Lunch	Free or reduced	16%
	Other	23%

Other findings from disaggregation on a longitudinal basis included:

1) a statistically significant reversal in the position of female students mathematics scores for the 11th and 12th grade cohorts over the three year period. That is, female students who exited grade 9 scoring nearly a standard deviation above males tended to emerge three years later nearly a standard deviation below their male counterparts (Sex of dropouts controlled).
2) the effects of tracking students by curriculum showed that there was almost no movement between and among the tracks over a three year

period. In addition, the relative proportion of students in the general/vocational track was consistently greater below the 30th percentile and thus highly correlated with social class.

In sum, the profile of this rural high school which emerged from a longitudinal analysis of the data was even less encouraging to teachers with respect to achievement than the snapshot would have been, but probably more realistic. For example, in interviews following the analysis, teachers suspected that grades 9 and 10 were currently performing better on tests than grades 11 and 12 had performed three years ago, but they also felt that low income students were not sharing in these gains. Overall gains, they thought, might be due to changes in curriculum and organization at the junior high level or might be an artifact of changes in the testing procedures in 1983 (an often overlooked threat to the validity of holding grade constant). Teachers also identified examples of sex stereotyping and track sterotyping which might account for patterns of achievement observed in those analyses.

In any case, there was considerably more to talk about, to ponder and potentially to act upon as a result of a more detailed and longitudinal analysis. Yet, this attempt raises questions for further study:

1) the procedure for identifying the grouping variables in this study consisted of asking the planning committee to list those variables by which they wanted to see the data disaggregated. This process usually occurs at the same time as the planners are reviewing other means of gathering data on school characteristics. What would happen if we waited until the data on school characteristics were gathered, processed and shared and then identified the grouping variables for achievement? Would the list be different? Should there be a standard list of grouping variables drawn from the research on school effectiveness? How many variables are enough or too many?
2) In identifying patterns of achievement, how many years are sufficient

in order to establish patterns which are meaningful to school planners?

3) In this study two measures of SES were used to describe the distributions, free and reduced lunch vs other and the counselors estimate of family social status (low vs other). While the former is in common usage in Michigan and Connecticut its sources of error include misplacement due to coding errors; and, more serious, mis-identification as a strategy for maintaining the free or reduced lunch child count. Counselor identification, thought by school personnel to be a more reliable indicator and conservative, was only modestly correlated with free lunch, (contingency coefficient of .65). This discrepancy seemed to affect the distribution proportions only slightly when achievement data were disaggregated (see Buttram and Carlson, 1983). What is the most appropriate measure of SES?

4) In order to avoid identifying teacher competence as a variable we currently avoid disaggregation on the basis of cohorts of students who constitute classroom groupings. Yet, this is a feasible strategy. Would teachers consent to such an analysis?

5) Descriptive statistics may be the most appropriate presentation of data which is non perscriptive during the planning process. On the other hand, at what point should more complicated time series analysis, covariance and discriminant analysis be employed? How would these findings be best presented to school planners?

References

Buttram, Joan and Carlson, Robert. "Effective School Research: Will It Play in the Country." Paper presented at AERA, April 11, 1983, Montreal, Canada.

Buttram, Joan and Carlson, Robert. "A Comparison of School Assessment Systems". Paper presented at AERA, April 27, 1984 New Orleans, La.

Buttram, Joan and Carlson, Robert. "The Implications of Effective Schools Research to the National Study of School Evaluation: Final Report." In process, 1984.

Edmonds, R.R. "Programs of School Improvement: An Overview." Educational Leadership, December, 1982.

Edmonds, R.R. "A Discussion of the Literature and Issues Related to Effective Schooling." Paper prepared for the National Conference on Urban Education, July 10-14, 1978, CEMREL, St. Louis, Mo.

Edmonds, R.R. "Testing and Educational Equity: The Status Quo and Prospects for the Future." in New Directions for Testing and Measurement: Measurement, Guidance, and Program Improvement, No. 13, San Francisco: Jossey-Bass, March, 1982.

Eubanks, E. and Levine, D. "A First Look at Effective School Projects at Inner City Elementary Schools." Paper presented at the American Association of School Administrators, Atlantic City, February 27, 1983.

Farrar, Eleanor, et al. "Volume III: Effective Schools Programs in High Schools: Implications for Policy, Practice and Research." Review of Effective Schools Programs. Cambridge, Ma: The Huron Institute, April, 1983.

Gauthier, Wm., et al. The Connecticut School Effectiveness Project: Development and Assessment. Hartford: The Connecticut State Department of Education, 1981.

Klitgaard, Robert and Hall, George. A Statistical Search for Unusually Effective Schools. Santa Monica: Rand Corproation, R-1210-CC/RC/ March, 1973.

Madaus, G.F., et al. School Effectiveness: A Reassessment of the Evidence. New York: McGraw-Hill, 1980.

McCormack, Maureen, et al. "Milwaukee's Project Rise." Educational Leadership, December, 1982.

Miles, Matthew, et al. "Volume II: The Extent of Adoption of Effective Schools Programs." Review of Effective Schools Programs. Cambridge, Ma: The Huron Institute. January, 1983.

Murnane, Richard. "Interpreting the Evidence on School Effectiveness." Working Paper No. 830. New Haven: Institution for Social and Policy Studies, December, 1980.

Neufield, Barbara, et al. "The Message for Secondary Schools." A Review of Effective Schools Research. Cambridge, Ma: The Huron Institute, January, 1983.

Oratz, Roz and McCarthy, Dennis. "An Overview of the Needs Assessment for a Single School." New York City Public Schools, School Improvement Project. New York: New York City Public Schools, 1981.

Pecheone, Ray and Shoemaker, Joan. "The Use of Multiple Data Sources to Assess School Effectiveness." Paper presented at AERA, March 20, 1982.